Use-conditional meaning in Rational Speech Act models[1]

Ciyang QING — *Stanford University*
Reuben COHN-GORDON — *Stanford University*

Abstract. Motivated by shortcomings in the modeling of repeated utterances in Burnett's *Rational Speech Act* (RSA) model of social meaning, we propose an alternative analysis in terms of use-conditional meaning, which we implement in an extended RSA framework. We show that it can not only model the production of a sequence of utterances throughout a discourse, but also captures the similarities between social and expressive meaning.

Keywords: social meaning, use-conditional meaning, Rational Speech Act models.

1. Introduction

Since Labov (1963), *variationist sociolinguistics* has studied the subtle meaning differences between linguistic variants (see Eckert, 2012 for an overview).

(1) a. I am walking (velar *-ing* variant)
b. I am walkin' (apical *-in'* variant)

For example, (1a) and (1b) have the same truth conditions, but after hearing them, listeners tend to infer different properties of the speaker: (1a) is associated with education, intelligence, articulateness (clustered as **competent**), and formality and distance (clustered as **aloof**) whereas (1b) is associated with the opposite properties (Campbell-Kibler, 2006, 2007, 2008, 2009). To the extent that the inference patterns are different, we can say that the variants in (1) have different *social meanings*. Moreover, people's production of language is sensitive to such social meanings and the context. For instance, Labov (2012) analyzes President Obama's use of (ING) in three recordings taken in three contexts with different levels of formality and finds 72% *-in'* at a barbecue, 33% at the press event that followed and 3% in his DNC acceptance speech.

These rich sociolinguistic phenomena have the potential to connect to semantics and pragmatics. Two fundamental research questions are (i) how to represent social meaning in the semantics and (ii) how to integrate it into the pragmatics. Burnett (2017, 2019) pioneers the use of Bayesian Rational Speech Act (RSA) models (Frank and Goodman, 2012; Goodman and Frank, 2016) to address these questions. In this paper we explore the connections between social meaning and semantics/pragmatics further. In particular, we address the following questions.

1. Empirically, what properties does social meaning have that resemble those studied in the semantics/pragmatics literature?
2. Theoretically, how do we semantically represent social meaning to capture these properties and model people's use of social meaning in production and comprehension?

The rest of the paper is organized as follows: in section 2, we list empirical properties that social meaning shares with expressive meaning. In section 3, we review and assess Burnett's

[1]We would like to thank Heather Burnett, Judith Degen, Dan Lassiter, Chris Potts, Penny Eckert, the anonymous reviewers and the participants of the SuB conference, for their very helpful comments and suggestions. All errors are our own.

Proceedings of Sinn und Bedeutung 23, vol. 2, pp. 253–265. Universitat Autònoma de Barcelona, Bellaterra (Cerdanyola del Vallès).

(2017, 2019) analysis of social meaning in the standard RSA framework, focusing on an issue that it has in modeling the production of a sequence of utterances throughout a discourse. In section 4, we propose an alternative analysis of social meaning in the spirit of Kaplan's (1999) use-conditional meaning, implement it in an extended RSA framework, and show that it can not only model the production of a sequence of utterances throughout a discourse, but can also capture the similarities between social and expressive meaning. We then discuss how this enriched RSA framework can serve as a general paradigm for other use-conditional meanings in section 5.

2. Social meaning and expressive meaning

Potts (2007) summarizes a list of empirical properties of *expressive meanings* (2).[2]

(2) Expressive meanings are

- largely independent of the descriptive meanings (Independence)
- always about the utterance situation itself. (Nondisplaceability)
- not propositional and can be hard to pin down (Descriptive ineffability)
- performative in that the very act of utterance conveys the meaning (Immediacy)
- strengthened when repeated without redundancy (Repeatability)

We note that social meaning also shares these properties. As a concrete example, consider the social meanings of the *-in'* variant discussed in section 1.

(3) John likes walkin'
$\rightsquigarrow$ The speaker is **incompetent**
$\rightsquigarrow$ The speaker is **friendly**

Let us check whether such social meanings also have the properties in (2):

- Independence: The social meanings of (3), which are about its speaker, are indeed independent of its descriptive meaning (i.e., the proposition that John likes walking).
- Nondisplaceability: The social meanings of *-in'* are about the speaker of the utterance, and therefore about the utterance situation itself. For instance, the social meanings of (4) are about its speaker, even though the *-in'* variant is embedded under past tense and *said*.

(4) Mary said that John was walkin'.
$\rightsquigarrow$ The speaker is **incompetent**
$\rightsquigarrow$ The speaker is **friendly**

- Descriptive ineffability: The social meaning of *-in'* is hard to pin down. The propositions in (3) are only approximations.
- Immediacy: Social meanings are performative. It is the very act of using the *-in'* variant that conveys its social meaning.

[2]For simplicity, we do not discuss the property of *perspective dependence* in Potts's list, but note that his treatment of it can be similarly incorporated in our proposal.

- Repeatability: the repeated use of the *-in'* variant can strengthen its social meaning without redundancy. For instance, the inferences in (5) can be stronger than in (3).

(5) John likes walkin'. Mary likes runnin'. Bob likes swimmin'.
$\rightsquigarrow$ The speaker is **incompetent**
$\rightsquigarrow$ The speaker is **friendly**

In sum, the empirical properties of expressive meanings in (2) are also shared by social meanings. These are properties that need to be accounted for by our analysis of social meaning.

3. Burnett's (2017, 2019) analysis of social meaning in the standard RSA framework

3.1. The Rational Speech Act (RSA) framework

Burnett's analysis is based on a Bayesian game-theoretic pragmatics framework, which provides a probabilistic formalization of Gricean reasoning (Frank and Goodman 2012; Goodman and Frank 2016; see also Franke 2009). In this framework, listeners and speakers recursively reason about each other (6).

(6) a. $L_0(w \mid u) \propto \Pr(w) \cdot [\![u]\!](w)$ Literal listener
b. $S_1(u \mid w) \propto \textbf{Optimize}(L_0(w \mid u))$ Pragmatic speaker
c. $L_1(w \mid u) \propto \Pr(w) \cdot S_1(u \mid w)$ Pragmatic listener

We will use a concrete example discussed by Burnett (2017) to illustrate the basics of this framework. Suppose you left three cookies on the dinner table for your roommates. When you come home, John, one of the roommates, tells you (7).

(7) I ate some of the cookies.

You will likely infer from (7) that John ate one or two cookies but probably not all three. An intuitive explanation is that if he had eaten all three cookies he would have told you (8), which is more informative.[3] How do we formalize this reasoning in the probabilistic framework above?

(8) I ate all of the cookies.

First, we consider a literal listener (6a), who interprets an utterance *u* simply based on its truth-conditional content $[\![u]\!]$. This can be seen as a probabilistic generalization of a Stalnakerian model of conversation: instead of treating a *context* *C* as a set of possible worlds, a context is now treated as a probability distribution $P(\cdot)$ over possible worlds. Updating an initial context (implemented as a *prior distribution* $\Pr(\cdot)$) with an utterance *u* results in a new context, i.e., the conditional probability $L_0(\cdot \mid u)$, obtained by probabilistic *conditioning* (9b) (cf. the traditional contextual update via set intersection (9a)).

(9) a. $C + u = C \cap [\![u]\!]$
b. $L_0(w \mid u) \propto \Pr(w) \cdot [\![u]\!](w)$

Concretely, in the cookie example, let w_i be the possible world in which John ate exactly i cookies ($i = 0, 1, 2, 3$). Just for illustration let us assume that the prior probabilities of w_0 to w_4 are 0.1, 0.4, 0.3, and 0.2, respectively (first row in (10)).

[3] Assume that it is clear that eating all three cookies is totally acceptable, so John would have no reason to hide the fact if he did eat them all.

(10) Literal listener after hearing *I ate some of the cookies*

	w_0	w_1	w_2	w_3
$\Pr(w)$	.1	.4	.3	.2
$[\![\text{some}]\!](w)$	0	1	1	1
$\Pr(w) \cdot [\![\text{some}]\!](w)$	0	.4	.3	.2
$L_0(w \mid \text{some})$	0	$\frac{4}{9}$	$\frac{3}{9}$	$\frac{2}{9}$

The utterance *I ate some of the cookies* is true in w_1, w_2, and w_3, i.e., $[\![\text{some}]\!](w_0) = 0$ and $[\![\text{some}]\!](w_i) = 1$ for $i = 1,2,3$ (second row in (10)). Based on the prior distribution $\Pr(\cdot)$ and the truth-conditional meaning $[\![\text{some}]\!]$, we compute their product in the third row in (10). The proportionality operator $\propto$ in (9b) means that its left term is obtained by multiplying a constant to the right term to ensure that the left term is a probability distribution (i.e., all the probabilities sum up to 1). This is done by dividing each element in the third row of (10) by the sum of that row, and the resulting conditional probabilities $L_0(w \mid \text{some})$ are shown in the last row. Note that $L_0(w_3 \mid \text{some}) = 2/9 \approx .22$, which means that after hearing *I ate some of the cookies*, the literal listener thinks it is roughly 22% likely that John ate all 3 cookies.

Similarly, the conditional probabilities $L_0(w \mid \text{all})$ are computed in (11).

(11) Literal listener after hearing *I ate all of the cookies*

	w_0	w_1	w_2	w_3
$\Pr(w)$	.1	.4	.3	.2
$[\![\text{all}]\!](w)$	0	0	0	1
$\Pr(w) \cdot [\![\text{all}]\!](w)$	0	0	0	.2
$L_0(w \mid \text{all})$	0	0	0	1

Note that $L_0(w_3 \mid \text{all}) = 1$, which means that after hearing *I ate all of the cookies*, the literal listener is completely sure that John ate all 3 cookies.

Now we will consider a pragmatic speaker, whose definition (6b) is repeated below as (12) to formalize the reasoning "if John had eaten all three cookies, he would have used *all* instead of *some* because *all* is more informative."

(12) $S_1(u \mid w) \propto \textbf{Optimize}(L_0(w \mid u))$ Pragmatic speaker[4]

What (12) says is that the probability of a pragmatic speaker choosing utterance u in world w is determined by optimizing (in some sense) the probability the literal listener assigns to w after hearing u. For the purposes of the current paper we only require that **Optimize** satisfies (13), i.e., the pragmatic speaker would be more likely to choose u than u' in world w iff the literal listener would assign higher probability to world w after hearing u than u'.[5]

(13) $S_1(u \mid w) > S_1(u' \mid w)$ iff $L_0(w \mid u) > L_0(w \mid u')$

[4]Here we make the simplifying assumption that all relevant utterances are equally likely *a priori*. A more realistic speaker model that takes into account utterance priors would be $S_1(u \mid w) \propto \Pr(u) \cdot \textbf{Optimize}(L_0(w \mid u))$. The simplification made here does not affect the main points of this paper.

[5]In many RSA models *a priori* preferences for utterances are represented as *costs* and the **Optimize** function is sensitive to them. In such cases (13) does not hold in general, but it still holds when all relevant utterances are assumed to be equally likely *a priori*.

In the cookie example, given that $L_0(w_3 \mid \text{all}) = 1 > L_0(w_3 \mid \text{some}) = 2/9$, from (10) we have $S_1(\text{all} \mid w_3) > S_1(\text{some} \mid w_3)$, i.e., the speaker would prefer to use *all* to *some* if he had eaten all 3 cookies.

Once we choose an exact definition of **Optimize**, we can make quantitative predictions about the production probabilities and apply Bayes' rule to make quantitative predictions about how a pragmatic listener would interpret an utterance (6c).

3.2. Personae and Eckert-Montague fields

In order to apply the RSA framework to analyze social meaning, we need to address two issues: (i) What do possible worlds represent? (ii) How do we represent social meaning?

Burnett (2019) uses Obama's use of (ING) as a working example and assume two candidate utterances $u_{\text{-}ing}$ and $u_{\text{-}in'}$ that are different only in terms of the realization of (ING).

For (i), Burnett uses possible worlds to represent the speaker's possible *personae*. A persona i is a maximally compatible set of properties. In the Obama example, Burnett assumes that the relevant properties are **competent** and **aloof** and their opposites. The four possible personae and their names are in the table below (14).

(14)

STERN LEADER	COOL GUY	ASSHOLE	DOOFUS
{**comp.**, **aloof**}	{**comp.**, **friendly**}	{**incomp.**, **aloof**}	{**incomp.**, **friendly**}

For (ii), Burnett essentially treats an utterance u's association with certain properties as a compatibility relation and uses type lifting to derive the set of personae that u is compatible with (i.e., treating personae as Montagovian individuals). A persona is compatible with u iff it contains a property that is compatible with u. Given that $u_{\text{-}ing}$ is compatible with both **competent** and **aloof**, the only persona incompatible with $u_{\text{-}ing}$ is DOOFUS. Similarly, since $u_{\text{-}ing}$ is compatible with both **incompetent** and **friendly**, the only persona incompatible with $u_{\text{-}in'}$ is STERN LEADER. Burnett calls such compatibility relations between linguistic variants and personae *Eckert-Montague fields* and uses them in the literal listener model as representations of social meaning (15).

(15)

	STERN LEADER	COOL GUY	ASSHOLE	DOOFUS
⟦*-ing*⟧	1	1	1	0
⟦*-in'*⟧	0	1	1	1

As a concrete example, Burnett models Obama's use of (ING) at the barbecue as follows. She assumes that, because Obama is the president, the prior probability distribution slightly favors personae that are **aloof** (i.e., STERN LEADER and ASSHOLE) than those that are **friendly** (i.e., COOL GUY and DOOFUS) (first row in (16)). Given the Eckert-Montague field of the *-ing* variant (second row in (16)), the probability that the literal listener assigns to each persona after hearing the *-ing* variant can be calculated according to (6a) and is shown in the last row in (16). In particular, $L_0(\text{COOL GUY} \mid \text{-}ing) = 2/8 = .25$. That is, after hearing the *-ing* variant, the literal listener thinks that it is 25% likely that Obama is a COOL GUY.

(16) Literal listener after hearing *-ing*

Persona i	STERN LEADER	COOL GUY	ASSHOLE	DOOFUS
$\Pr(i)$	.3	.2	.3	.2
$[\![\textit{-ing}]\!](i)$	1	1	1	0
$L_0(i \mid \textit{-ing})$	$\frac{3}{8}$	$\frac{2}{8}$	$\frac{3}{8}$	0

Similarly, the probability that the literal listener assigns to each persona after hearing the *-in'* variant is calculated in (17). In particular, $L_0(\text{COOL GUY} \mid \textit{-in'}) = 2/7 \approx .286$.

(17) Literal listener after hearing *-in'*

Persona i	STERN LEADER	COOL GUY	ASSHOLE	DOOFUS
$\Pr(i)$	.3	.2	.3	.2
$[\![\textit{-in'}]\!](i)$	0	1	1	1
$L_0(i \mid \textit{-in'})$	0	$\frac{2}{7}$	$\frac{3}{7}$	$\frac{2}{7}$

Given that $L_0(\text{COOL GUY} \mid \textit{-in'}) > L_0(\text{COOL GUY} \mid \textit{-ing})$, from (13) we can conclude that $S_1(\textit{-in'} \mid \text{COOL GUY}) > S_1(\textit{-ing} \mid \text{COOL GUY})$, i.e., assuming that Obama wants to convey the COOL GUY persona, he would prefer the *-in'* variant to the *-ing* variant. Recall that in this context Obama in fact uses *-in'* 72% of the time. Therefore, it seems that Burnett's model correctly captures Obama's use of (ING) in this case. (In fact, using a standard implementation of the **Optimize** function, she is able to configure the model so that $S_1(\textit{-in'} \mid \text{COOL GUY})$ is predicted to be around 69%.)

3.3. The issue with multiple utterances

However, when we examine more closely Burnett's model's prediction $S_1(\textit{-in'} \mid \text{COOL GUY})$, it is not clear whether this truly captures Obama's production rate of *-in'*.

Crucially, note that $S_1(\textit{-in'} \mid \text{COOL GUY})$ is implicitly relative to the literal listener's prior probability distribution over personae. This gives rise to the important question of how such a probability distribution should change over the entire discourse, which may consist of multiple utterances in a sequence. To address this, recall that the literal listener model can be seen as a probabilistic generalization of a Stalnakerian model (9), repeated below as (18).

(18) a. $C + u = C \cap [\![u]\!]$
b. $L_0(i \mid u) \propto \Pr(i) \cdot [\![u]\!](i)$

If we have a second utterance u_2 following the first u_1, in a Stalnakerian model we can take the output context $C' = C + u_1$ as the input context of u_2. By analogy, we can use the conditional probability $L_0(w \mid u_1)$ as the new prior probability $\Pr'(w)$ in the literal listener model (19).

(19) $L_0'(i \mid u_2) \propto \Pr'(i) \cdot [\![u_2]\!](i)$

In general, for a sequence of utterances $u_1, \ldots, u_{n+1}$, we can define the literal listener model recursively as in (20).

(20) $L_0^{(n)}(i \mid u_1, \ldots u_{n+1}) \propto \Pr^{(n)}(i) \cdot [\![u_{n+1}]\!](i), \quad \Pr^{(n+1)}(w) = L_0^{(n)}(i \mid u_1, \ldots u_{n+1})$

As notational variants we will write $L_0^{(0)}, \Pr^{(0)}$ simply as $L_0, \Pr$, write $L_0^{(1)}, \Pr^{(1)}$ as $L_0', \Pr'$, write $L_0^{(2)}, \Pr^{(2)}$ as $L_0'', \Pr''$, etc. Also, when the initial sequence $u_1, \ldots, u_n$ is clear from the context, we will write $L_0^{(n)}(i \mid u_{n+1})$ instead of $L_0^{(n)}(i \mid u_1, \dots u_{n+1})$

Once the literal listener is defined over a sequence of utterances, the pragmatic speaker model can be similarly generalized (21).

(21) $S_1^{(n)}(u_{n+1} \mid i) \propto \textbf{Optimize}(L_0^{(n)}(i \mid u_{n+1}))$

Now let us consider what Burnett's model predicts when the speaker needs to make multiple utterances, again using Obama's use of (ING) as a concrete example.

From the previous discussion we already know that S_1(*-in'* | COOL GUY) > S_1(*-ing* | COOL GUY) and therefore let us assume that Obama first chooses the *-in'* variant and see what the model predicts about his second utterance.

In this case, the probability that the literal listener assigns to each persona after hearing *-ing* as the second utterance is computed in (22). Note that the prior $\Pr'(i)$ is by definition L_0(i | *-in'*), i.e., the last row in (17).

(22) Literal listener after first hearing *-in'* and then *-ing*

Persona i	STERN LEADER	COOL GUY	ASSHOLE	DOOFUS
$\Pr'(i)$	0	$\frac{2}{7}$	$\frac{3}{7}$	$\frac{2}{7}$
⟦-ing⟧(i)	1	1	1	0
L'_0(i | *-ing*)	0	$\frac{2}{5}$	$\frac{3}{5}$	0

Similarly, the probability that the literal listener assigns to each persona after hearing *-in'* as the second utterance is computed in (23). Note that the conditional probability L'_0(i | *-in'*) is identical to the prior $\Pr'(i)$: Since all the personae with non-zero prior probabilities are already compatible with the *-in'* variant, no information is gained from hearing it.

(23) Literal listener after first hearing *-in'* and then *-in'* again.

Persona i	STERN LEADER	COOL GUY	ASSHOLE	DOOFUS
$\Pr'(i)$	0	$\frac{2}{7}$	$\frac{3}{7}$	$\frac{2}{7}$
⟦-in'⟧(i)	0	1	1	1
L'_0(i | *-in'*)	0	$\frac{2}{7}$	$\frac{3}{7}$	$\frac{2}{7}$

Since we have L'_0(COOL GUY | *-ing*) $= \frac{2}{5} >$ L'_0(COOL GUY | *-in'*) $= \frac{2}{7}$, from (13) we conclude that S'_1(*-ing* | COOL GUY) > S'_1(*-in'* | COOL GUY). That is, after using the *-in'* variant once, it is no longer preferred, because it is not informative anymore.

Finally, let us assume that Obama first uses *-in'*, and then *-ing*, and see what Burnett's model predicts about the third utterance.

In this case, the probability that the literal listener assigns to each persona after hearing *-ing* as the third utterance is computed in (24). Note that the prior $\Pr''(i)$ by definition comes from the last row in (22). Also note that the conditional probability L''_0(i | *-ing*) is identical to the prior $\Pr''(i)$, because all the persona with non-zero priors are already compatible with *-ing*.

(24) Literal listener after first hearing *-in'*, followed by *-ing*, and then *-ing*

Persona i	STERN LEADER	COOL GUY	ASSHOLE	DOOFUS
$\Pr''(i)$	0	$\frac{2}{5}$	$\frac{3}{5}$	0
$[\![\text{-ing}]\!](i)$	1	1	1	0
$L''_0(i \mid \textit{-ing})$	0	$\frac{2}{5}$	$\frac{3}{5}$	0

Similarly, the probability that the literal listener assigns to each persona after hearing *-in'* as the third utterance is computed in (25). Once again, note that the conditional probability $L''_0(i \mid \textit{-in'})$ is also identical to the prior $\Pr''(i)$.

(25) Literal listener after first hearing *-in'*, followed by *-ing*, and then *-in'*

Persona i	STERN LEADER	COOL GUY	ASSHOLE	DOOFUS
$\Pr''(i)$	0	$\frac{2}{5}$	$\frac{3}{5}$	0
$[\![\text{-in'}]\!](i)$	0	1	1	1
$L''_0(i \mid \textit{-in'})$	0	$\frac{2}{5}$	$\frac{3}{5}$	0

Since we have $L''_0(\text{COOL GUY} \mid \textit{-ing}) = \frac{2}{5} = L''_0(\text{COOL GUY} \mid \textit{-in'})$, from (13) we conclude that $S''_1(\textit{-ing} \mid \text{COOL GUY}) = S''_1(\textit{-in'} \mid \text{COOL GUY})$. That is, after using both *-in'* and *-ing* variants once, Burnett's model predicts that the speaker should have no preference for either variant.

In sum, Burnett's model predicts that at the barbecue (i) Obama would initially prefer the *-in'* variant, (ii) after he has produced one variant but not both, he would prefer the one he has not produced, and (iii) after he has produced both variants, he would be indifferent. Together, this means that Obama's production rate of *-in'* is predicted to be around 50% when the speech is sufficiently long (i.e., there are enough instances of (ING) after he produces both variants), contrary to fact.

Note that the problem with multiple utterances does not crucially rely on the way the sequential update of the prior distribution is defined in (20). For instance, after hearing the first *-in'*, we can compute the pragmatic listener's interpretation $L_1(i \mid \textit{-in'})$ and use it as the new prior $\Pr'(i)$, instead of using $L_0(i \mid \textit{-in'})$ as $\Pr'(i)$. However, the same problem will arise, because it is still the case that every persona i whose new prior $\Pr'(i)$ is greater than 0 is already compatible with *-in'*. This means that *-in'* is totally uninformative (i.e., $L'_0(i \mid \textit{-in'})$ is identical to $\Pr'(i)$) and therefore will never be preferred.

However, we note that this problem does depend on the simplifying assumption that the persona intended by the speaker stays constant throughout the sequence of utterances in the discourse. If we give up this assumption, and analyze the social meaning of a linguistic variant as a proposition relativized to the utterance time t (e.g., the meaning of *-ing* is "I am competent or aloof *now*"), then the repeated use of a variant can be informative because two instances of use correspond to different propositions (e.g., "I am competent or aloof at t_1/t_2").

Although such an approach can potentially rescue Burnett's analysis, many more complicated issues need to be addressed to provide a realistic model. For instance, in order to calculate the speaker's production rate of a variant over a sequence of utterances in the discourse, we will need to specify how the prior over personae is updated throughout the discourse, which is more complicated because we need to take into account the potential change in persona.

Therefore, we conclude that Burnett's analysis faces difficulties in modeling the production of a sequence of utterances throughout the discourse. In the next section, we propose an alternative analysis of social meaning in terms of use-conditional meaning, and show that it not only can easily model the production of a sequence of utterances throughout the discourse, but also captures the similarities between social meaning and expressive meaning very well.

4. Social meaning as use-conditional meaning

4.1. Use-conditional meaning

According to Kaplan (1999), the meanings of expressions such as *ouch* and *oops* are not captured by their truth conditions (since they do not have any), but rather by their conditions on use, i.e., contexts in which they can be correctly/felicitously used. For instance, the use condition of *oops* can be roughly characterized as follows: *oops* is felicitously used iff the speaker observes a minor mishap. Knowing this use condition, the listener of *oops* can infer that the speaker observes a minor mishap, by virtue of knowing what it takes for the speaker to produce this utterance.

In the case of social meaning on the one hand, the choice of variant is not totally constrained by linguistic norms. For instance, anybody may use *-ing* and *-in'* and using either variant would be correct. On the other hand, we can still apply Kaplan's insight to analyze social meaning, in that the hearer of a linguistic variant can gain information about the speaker, by virtue of knowing (or more precisely, having an ideology about) what it takes for the speaker to produce that variant, which often has to do with the speaker's social identity, or *persona*.

Therefore, we suggest that social meaning should be represented as the listener's ideology about how the speaker's persona influences his/her choice of the linguistic variants. Below we formalize this idea in an extended RSA framework.

4.2. Implementing use-conditional meaning in an extended RSA framework

We generalize the literal listener to integrate both the truth-conditional (26a) and use-conditional (26b) meanings of an utterance.

(26) $L_0(w,i \mid u) = L_0(w \mid u) \cdot L_0(i \mid u)$

a. $L_0(w \mid u) \propto \Pr(w) \cdot [\![u]\!](w)$ — $[\![u]\!]$: truth-conditional meaning

b. $L_0(i \mid u) \propto \Pr(i) \cdot S_0(u \mid i)$
where $S_0(u \mid i)$ is a *stereotypical speaker* — $S_0(u \mid i)$: use-conditional meaning

The literal listener updates their prior belief about the world w by conditioning on the truth of the utterance u (26a). Meanwhile, they update their prior belief about the speaker type i by reasoning about a hypothetical *stereotypical speaker* S_0.[6] Crucially, this speaker is *not* a production model of any real linguistic agent, but rather one that is based on ideological stereotypes. For instance, the stereotype that incompetent but friendly people tend to use the *-in'* variant corresponds to a high value for $S_0(\textit{-in'} \mid \text{DOOFUS})$, while the stereotype that competent but aloof people tend to use the *-ing* variant corresponds to a low value of $S_0(\textit{-in'} \mid \text{DOOFUS})$. Assuming that the production probabilities of the other two personae are in between, we take (27) as an example representation of the social meaning of *-in'* and use it to model Obama's use of (ING) at the barbecue.

[6]See Henderson and McCready's (2017) analysis of dogwhistles for another case of using a speaker model S_0 as the starting point of the iterative reasoning.

(27)

i	STERN LEADER	COOL GUY	ASSHOLE	DOOFUS
$S_0(\textit{-in'} \mid i)$	0.1	0.7	0.2	0.9

4.3. Multiple utterances

First, we consider Obama's first utterance. Assuming the same prior probability distribution over personae as before, which slightly favors personae that are **aloof**, the probability that the literal listener assigns to each persona after hearing *-in'* is computed in (28).

(28) Literal listener after hearing *-in'*

i	STERN LEADER	COOL GUY	ASSHOLE	DOOFUS
$\Pr(i)$	0.3	0.2	0.3	0.2
$S_0(\textit{-in'} \mid i)$	0.1	0.7	0.2	0.9
$\Pr(i) \cdot S_0(\textit{-in'} \mid i)$	0.03	0.14	0.06	0.18
$L_0(i \mid \textit{-in'})$	0.073	0.34	0.146	0.439

Similarly, the probability that the literal listener assigns to each persona after hearing *-ing* is computed in (29). Note that $S_0(\textit{-ing} \mid i) = 1 - S_0(\textit{-in'} \mid i)$ for each persona i.

(29) Literal listener after hearing *-ing*

i	STERN LEADER	COOL GUY	ASSHOLE	DOOFUS
$\Pr(i)$	0.3	0.2	0.3	0.2
$S_0(\textit{-ing} \mid i)$	0.9	0.3	0.8	0.1
$\Pr(i) \cdot S_0(\textit{-ing} \mid i)$	0.27	0.06	0.24	0.02
$L_0(i \mid \textit{-ing})$	0.458	0.102	0.407	0.034

Since $L_0(\text{COOL GUY} \mid \textit{-in'}) = 0.34 > 0.102 = L_0(\text{COOL GUY} \mid \textit{-ing})$, from (13) we conclude that $S_1(\textit{-in'} \mid \text{COOL GUY}) > S_1(\textit{-ing} \mid \text{COOL GUY})$. That is, Obama would initially prefer *-in'* to *-ing* to construct the COOL GUY persona.

Now suppose Obama chooses *-in'* as the first utterance. The probabilities that the literal listener assigns to each persona after hearing *-in'* or *-ing* as the second utterance are shown in (30).

(30) Literal listener after hearing *-in'* first, and then *-in'* or *-ing* ($\Pr'(i) = L_0(i \mid \textit{-in'})$)

i	STERN LEADER	COOL GUY	ASSHOLE	DOOFUS
$\Pr'(i)$	0.073	0.34	0.146	0.439
$S_0(\textit{-in'} \mid i)$	0.1	0.7	0.2	0.9
$L'_0(i \mid \textit{-in'})$	0.011	0.356	0.044	0.589
$L'_0(i \mid \textit{-ing})$	0.2	0.311	0.356	0.133

Since $L'_0(\text{COOL GUY} \mid \textit{-in'}) = 0.356 > 0.311 = L'_0(\text{COOL GUY} \mid \textit{-ing})$, from (13) we conclude that $S'_1(\textit{-in'} \mid \text{COOL GUY}) > S'_1(\textit{-ing} \mid \text{COOL GUY})$. That is, after using the *-in'* variant once, Obama would still prefer it to *-ing* as the second utterance to construct the COOL GUY persona. This prediction is in stark contrast with Burnett's model's, according to which Obama would prefer *-ing* as the second utterance once he has used *-in'* as the first utterance. The reason is that in our model *-in'* never completely rules out any persona and therefore its repeated use can still be informative, whereas in Burnett's model, repeated use of *-in'* is never informative and therefore the repeated use is never preferred.

Now suppose Obama chooses *-in'* as the first and second utterances. The probabilities that the literal listener assigns to each persona after hearing *-in'* or *-ing* as the third utterance are shown in (31).

(31) Literal listener after hearing *-in'* twice, and then *-in'* or *-ing* ($\mathrm{Pr}''(i) = L'_0(i \mid \textit{-in'})$)

i	STERN LEADER	COOL GUY	ASSHOLE	DOOFUS
$\mathrm{Pr}''(i)$	0.011	0.356	0.044	0.589
$S_0(\textit{-in'} \mid i)$	0.1	0.7	0.2	0.9
$L''_0(i \mid \textit{-in'})$	0.001	0.316	0.011	0.672
$L''_0(i \mid \textit{-ing})$	0.047	0.508	0.166	0.28

Since $L''_0(\text{COOL GUY} \mid \textit{-in'}) = 0.316 < 0.508 = L''_0(\text{COOL GUY} \mid \textit{-ing})$, from (13) we conclude that $S''_1(\textit{-in'} \mid \text{COOL GUY}) < S''_1(\textit{-ing} \mid \text{COOL GUY})$. That is, after using *-in'* twice, Obama would prefer to use *-ing* as the third utterance to construct the COOL GUY persona.

This ensures that Obama will not use the *-in'* variant indefinitely. Intuitively, this is because using the *-in'* variant too much will make the listener assign too much probability to the DOOFUS persona, since that persona corresponds to the highest production rate of *-in'*.

In general, the speaker's best strategy to convey the intended persona i to the listener over multiple utterances is to produce *-in'* at the rate that conforms to the stereotype of that persona i, i.e., $S_0(\textit{-in'} \mid i)$. For example, Obama's best strategy to convey the COOL GUY persona is to produce the *-in'* variant 70% of the time. We suggest that this captures the performative nature of social meaning: the speaker can succeed in conveying an intended persona iff the way he/she produces the utterances conforms to the way the intended persona is supposed to produce those utterances. In other words, it is the very act of using linguistic variants in a particular way that constructs and conveys the corresponding persona.

4.4. Capturing properties of expressive meanings

As discussed earlier, social meaning shares many properties of expressive meanings (2), repeated below as (32).

(32) Expressive meanings are

- largely independent of the descriptive meanings (Independence)
- always about the utterance situation itself. (Nondisplaceability)
- not propositional and can be hard to pin down (Descriptive ineffability)
- performative in that the very act of utterance conveys the meaning (Immediacy)
- strengthened when repeated without redundancy (Repeatability)

We have discussed how our proposal captures Immediacy and Repeatability. The remaining properties are discussed below.

- Independence holds due to our assumption that w and i are conditionally independent given u, which is generally the case. However, sometimes this assumption might fail to hold. Then we use the most general formula $L_0(w,i \mid u) \propto \Pr(w,i) \cdot [\![u]\!](w) \cdot S_0(u \mid [\![u]\!], w, i)$, which allows for descriptive and indexical meanings to interact with each other.

- Nondisplaceability: Social meaning, represented as a production probability $S_0(u \mid i)$, is always about the utterance situation itself, since the persona i is always relative to the speaker of the utterance.
- Descriptive ineffability: social meaning is represented as a production probability $S_0(u \mid i)$, which is not propositional. It can be hard to pin down because it can be difficult to describe the high likelihood region(s) of this production probability.

5. Further discussion

Representing use-conditional meanings in terms of the stereotypical speaker S_0 is a natural extension to previous multi-dimensional semantic systems (e.g., Potts, 2007; Gutzmann, 2015).

As a concrete example, consider the following simplified reformulation of Potts's analysis of the expression *the damn dog*. This expression has a multi-dimensional meaning. On the one hand, it has a normal descriptive content **d** (a type *e* individual), which comes from [[the dog]]. On the other, it also has an expressive content f, a function that updates/alters the context. Here the context is represented by an *expressive index* of the form $\mathbf{s} I \mathbf{d}$, where I is an *expressive interval*, i.e., a subset of $[-1,1]$ representing the set of possible attitudes of the speaker **s** towards **d**. The effect of f is to shrink the expressive interval I and ensure that it is a subset of $[-1,0]$, i.e., the speaker **s** holds negative attitudes towards the dog **d**.

Generalizing the expressive interval I to a probability distribution $p(i)$ over the speaker's possible attitude i towards the dog **d**, we can derive a probabilistic literal listener model based on (26b), as shown in (33).

(33) $L_0(i \mid damn) \propto \Pr(i) \cdot S_0(damn \mid i)$

We can see that, once $S_0(damn \mid i)$ is specified, (33) will implicitly define a mapping from the input context $\Pr(i)$ to an updated context $L_0(i \mid damn)$, i.e., the expressive content of *the damn dog*. In other words, the expressive meaning of *the damn dog* can be represented as $S_0(damn \mid i)$, a probabilistic use-conditional meaning.

Note that this example can be generalized to a fully compositional analysis of *damn*: its descriptive content is the identity function **id**, and its expressive meaning is represented as $\lambda k.S_0(damn \mid i[k^d(\mathbf{id})])$. This expressive meaning has two main components. The first is $\lambda x.S_0(damn \mid i[x])$, the probabilistic use-conditional meaning in (33) relativized to x, i.e., how likely a speaker will use *damn*, given that his/her attitude towards x is $i[x]$. Here x can be an individual (e.g., the dog **d** in the above example), a proposition (e.g., the fact that the speaker lost the key in *I lost the damn key*), etc. The second component specifies that x is determined by the scope of *damn*. Concretely, x is the result of feeding the descriptive content of the expression that *damn* takes scope over with the identity function as its argument, i.e., $k^d(\mathbf{id})$. For instance, when *damn* takes scope over the rest of the DP in *the damn dog*, i.e., the descriptive content of the expression that *damn* takes scope over would be $\lambda f.\mathbf{the}(f(\mathbf{dog})$, feeding it with **id** will make x be $\mathbf{the}(\mathbf{dog})$ (i.e., the dog **d**). When *damn* takes scope over the rest of the sentence in *I lost the damn key* (schematically, damn(λf. I lost the (f key)), applying the same procedure makes x be the proposition that the speaker lost the key.

This probabilistic approach provides a principled way to concretely specify expressive meanings and explain their properties. For instance, when the expressive meaning of *damn* is represented as a function that shrinks expressive intervals, it can be hard and arbitrary to specify how much the function should shrink each interval. In contrast, when the expressive meaning is represented a production probability $S_0(damn \mid i)$, it is much easier to specify an intuitively plausible production probability or learn it from empirical data.

6. Conclusion

In this paper, we analyze social meaning as use-conditional meaning, and provide a formal analysis by extending the RSA framework. We show that such an analysis can well capture the properties social meaning shares with expressive meanings, and discuss how the extended probabilistic framework could model use-conditional meanings in general.

References

Burnett, H. (2017). Sociolinguistic interaction and identity construction: The view from game-theoretic pragmatics. *Journal of Sociolinguistics 21*(2), 238–271.

Burnett, H. (2019). Signalling games, sociolinguistic variation and the construction of style. *Linguistics and Philosophy*, 1–32.

Campbell-Kibler, K. (2006). *Listener perceptions of sociolinguistic variables: The case of (ING)*. Ph. D. thesis, Stanford University.

Campbell-Kibler, K. (2007). Accent, (ING), and the social logic of litsner perceptions. *American Speech 82*(1), 32–64.

Campbell-Kibler, K. (2008). I'll be the judge of that: Diversity in social perceptions of (ING). *Language in Society 37*(5), 637–659.

Campbell-Kibler, K. (2009). The nature of sociolinguistic perception. *Language Variation and Change 21*(1), 135–156.

Eckert, P. (2012). Three waves of variation study: The emergence of meaning in the study of sociolinguistic variation. *Annual review of Anthropology 41*, 87–100.

Frank, M. C. and N. D. Goodman (2012). Predicting pragmatic reasoning in language games. *Science 336*(6084), 998–998.

Franke, M. (2009). *Signal to Act: Game Theory in Pragmatics*. Ph. D. thesis, Institute for Logic, Language and Computation, Universiteit van Amsterdam.

Goodman, N. D. and M. C. Frank (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences 20*(11), 818–829.

Gutzmann, D. (2015). *Use-conditional meaning: Studies in Semantics and Pragmatics*. Oxford: Oxford University Press.

Henderson, R. and E. McCready (2017). Dog-whistles and the at-issue/not-at-issue distinction. Presentation at the DGfS Workshop: Secondary Information & Lingustic Encoding.

Kaplan, D. (1999). The meaning of *ouch* and *oops*: Explorations in the theory of meaning as use. Unpublished manuscript, University of California, Los Angeles.

Labov, W. (1963). The social motivation of a sound change. *Word 19*(3), 273–309.

Labov, W. (2012). *Dialect diversity in America: The politics of language change*. Charlottesville, VA: University of Virginia Press.

Potts, C. (2007). The expressive dimension. *Theoretical Linguistics 33*(2), 165–197.